ORGIA INSTITUTE OF TECHNOLOGY — OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

[X] ORIGINAL [ ] REVISION NO. ____

ject No. B-556 DATE 9-8-81

ject Director: K. R. Perry ~~School~~/Lab SEL

nsor: NASA - Ames University Consortium

pe Agreement: Interchange Agreement No. NCA2-OR260-101

ard Period: From 8-1-81 To 7-31-82 (Performance) 7-31-82 (Reports)

nsor Amount: $29,043; Partially funded at $14,950 Contracted through:

st Sharing: ~~GTRI~~/GIT

le: Digital Signal Processing for Cardiovascular Physiological Monitoring Applications

MINISTRATIVE DATA OCA Contact Dan S. Tasty

Sponsor Technical Contact:
Dr. Harold Fandler, Chief
Biomedical Research Division
Mail Stop 239-8
NASA-Ames Research Center
Moffett Field, California 94035
Phone (415) 965-5745

2) Sponsor Admin/Contractual Matters:
Capt John W. Hines, Jr
Biomedical Research Division
Mail Stop 236-6
NASA-Ames Research Center
Moffett Field, California 94035
Phone (415) 965-5741

fense Priority Rating: N/A Security Classification: N/A

STRICTIONS

Attached ____ Supplemental Information Sheet for Additional Requirements.

vel: Foreign travel must have prior approval – Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

uipment: Title vests with N/A - None Proposed

MMENTS:
This project is funded as Fixed Price.

SEP 1981 RECEIVED Research Reports Office

PIES TO:

- Administrative Coordinator
- Research Property Management
- Accounting
- Procurement/EES Supply Services
- ~~Reports Coordinator (OCA)~~
- Legal Services (OCA)
- Library
- EES Public Relations (2)
- Computer Input
- Project File
- Other ____

RM OCA 4-78)

ORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 5/1/84

ect No. B-556 School/Lab SEL

ides Subproject No.(s) N/A

ect Director(s) K. R. Perry GTRI / GIT

nsor NASA/Ames University Consortium

Digital Signal Processing For Cardiovascular Physiological Monitoring Applications

ctive Completion Date: 7/31/82 (Performance) 7/31/82 (Reports)
Fixed Price

nt/Contract Closeout Actions Remaining:

- [x] None
- [ ] Final Invoice or Final Fiscal Report
- [ ] Closing Documents
- [ ] Final Report of Inventions
- [ ] Govt. Property Inventory & Related Certificate
- [ ] Classified Material Certificate
- [ ] Other

tinues Project No. N/A Continued by Project No. N/A

IES TO:

ect Director
earch Administrative Network
earch Property Management
ounting
urement/EES Supply Services
earch Security Services
orts Coordinator (OCA)
al Services

Library
GTRI
Research Communications (2)
Project File
Other

n OCA 60:1028

B-5

Georgia Institute of Technology
ENGINEERING EXPERIMENT STATION
Atlanta, Georgia 30332

09 November 1983

Captain John W. Hines, Jr.
NASA Ames Research Center
Biomedical Research Division
Cardiovascular Research Laboratory
Moffett Field, CA 94035

Dear Capt. Hines:

Submitted herewith is the final report on the Joint Research Interchange, "Digital Signal Processing for the NASA Ames Cardiovascular Research Laboratory." Since this project was stopped midway through its completion, the other two deliverable reports will not be necessary. (The first progress report was submitted February 26, 1983.)

This final report contains five sections. The first section discusses the mean cancellation circuit. The second section discusses the potential application of adaptive equalization to in-vivo blood pressure measurements. The third section discusses the different data compression/data reduction algorithms that were investigated and presents the conclusion that future data compression/data reduction algorithms must contain a measure of diagnostic value, in conjunction with physiologists, for the comparison of the computerized diagnostic tests versus the resultant compressed waveform. And the fourth section presents a mathematical analysis of the segmented filter bank approach for doppler processing. The conclusion reached in this section is that if we divide the original signal bandwidth into N identical banks, we derive a square root on N improvement in the signal-to-noise ratio. The fifth section is an appendix to the second section. The appendix discusses the principles of adaptive signal processing.

I hope that this final report meets with your approval. If, however, you have any questions concerning its content, please call me at (404) 894-3513. And I would be more than happy to answer any questions that you may have or to discuss the possibilities of reviving and completing the original project.

Sincerely yours,

Kenneth R. Perry, Ph.D.
Research Engineer

KRP/sw

FINAL REPORT

DIGITAL SIGNAL PROCESSING FOR THE NASA AMES CARDIOVASCULAR RESEARCH LABORATORY

Submitted to

Captain John W. Hines, Jr.
NASA Ames Research Center
Biomedical Research Division
Cardiovascular Research Laboratory
Moffett Field, CA 94035

by

Georgia Institute of Technology
Engineering Experiment Station
Systems Engineering Laboratory
Atlanta, GA 30332

Contracting through

Georgia Tech Research Institute

11 October 1983

Digital Signal Processing for Cardiovascular Physiological Monitoring Applications

Final Report

I. Mean Cancellation

One of the difficulties with the use of the strain gauge transducers in implantable telemetry systems is the signal amplitude and the drift with respect to time. Because of the drift characteristics of the transducers the allowable signal amplitude is often reduced in order to permit extended usage of the devices. Investigators frequently desire more signal amplitude and dynamic range in measuring the parameter of interest. Engineers are faced with the difficulty of either increasing the amplitude of the signal at the cost of reduced system lifetime due to the drift characteristics or maintaining a more conservative output level in order to ascertain extended usage of the device.

Kelvin's method can be implemented to model a continuous time system which is described by a differential equation. The differential equation is modeled with integrators instead of differentiators. In reality, all signals are corrupted by some amount of noise. When such a signal is differentiated, more noise results. The amplitude of this noise is usually large enough to "drown out" the derivative of the signal, therefore, integrators are used instead of differentiators.

The basic concept is shown in Figure 1. The normal pulsatile output of the signal contains the pulsatile component, S1, the mean component, M1, and miscellaneous offset components, O1, due to drift and other artifacts. Extracting the mean

via a low pass integrator filter, inverting, and summing with S1 shifts S1 to the zero output baseline. Level shifting circuitry allows positioning of the signal to midline (S2). The shifted signal can now be amplified to provide more dynamic range of the original signal, S1, while still retaining the mean information, M1. The one disadvantage in this circuit is that it requires another output channel for data acquisition. In multichannel telemetry systems, techniques such as this can significantly increase the usable lifetime of the system, resulting in the ability to conduct longer chronic studies at reduced costs.

This technique has been successfully tested in the Cardiovascular Research Lab.

II. Application of Adaptive Equalization to In-Vivo Blood Pressure Measurements

Measurement of intracardiac pressures using a fluid-filled catheter is usually distorted by the hydraulic response characteristics of the transmission medium. The hydraulic system can be diagrammed as shown in Figure 2. The inlet of the pressure gauge is assumed to have a volume U, with the catheter assumed at length l. A lumped circuit analogy is described in figure 3 (Welkowitz and Deutsch, 1976). The analogous electrical circuit elements can be related to the physical properties as follows:

$$L = \rho_D \ell_e / U^2 \qquad R = 8\pi\ell/\pi U^4$$

$$C = U/\rho_D v_s^2 \qquad \text{WHERE: } \ell_e = \ell = 0.8(\pi r^2)^{\frac{1}{2}}$$

ρ_D = Fluid Density; λ = Fluid Viscosity

From the analogous network, the transfer function of the coupling system is given by:

$$F(j\omega) = Ps/Pc = \frac{1}{(1 - \omega^2 LC + j\omega RC)}$$

The magnitude response is:

$$|F(j\omega)| = |P_G/Pc| = \frac{1}{\left[(1-\omega^2 LC)^2 + \omega^2 C^2 R^2\right]^{\frac{1}{2}}}$$

Since the transfer function is known, the output of the system can be passed through the mathematical inverse filter of the measurement system. This should allow accurate reconstruction of the transmitted signal by cancellation of the effects of the hydraulic fluid-filled catheter. This technique should be implemented and compared with measurements made simultaneously at the point of interest.

Continued development toward a standard method for comparing direct versus indirect measurements of blood pressure will result in bioinstrumentation to automatically compensate, on-line and in real-time, for the distortion in blood pressure readings caused by the nonlinear frequency response of the fluid-filled catheter. This will allow researchers, for the first time, to directly compare invasive and noninvasive blood presure measurements.

The proposed system would implement an adaptive compensation network for instantaneous frequency response equalization of the fluid-filled tube/blood pressure transducer system. The adaptive system will provide a dynamic model of the

blood pressure measurement system. (See figure 4.) This time-variant model of the all-pole blood pressure measurement system can then be used to compensate for the distortion caused by the nonstationary, nonlinear frequency response of the fluid filled tube which conducts the pressure wave from the in-dwelling catheter to the pressure transducer.

Many investigations have compared the performance of noninvasive versus invasive measurements of blood pressure. However, none of these researchers had a method to compensate for the instantaneous changes in the transfer function of the transmission medium, the fluid-filled tube. A recent paper by J. Bruner, et al (Bruner, 1981), illustrates the problem in comparing direct versus indirect blood pressure measurements,

> "the frequency response of commercially available transducers far exceeds that needed for clinical investigations, but the addition of a fluid-filled tube between the transducer and the invaded vessel creates a resonant system. The performance of the resonant system is defined by the resonant frequency and the damping factor. All practical systems in clinical use are characterized by low resonant frequencies; all are underdamped... We suggest that system frequency response is at the root of measurement discrepencies between direct and indirect techniques, and accounts for much of the conflict between studies comparing direct and indirect measurements."

The design of this system involves the application of optimal and adaptive control theory techniques to a system defined in terms of classical control theory parameters. The extensive use of mathematics in optimal control theory has made it difficult to quickly translate between classical and modern control theory parameters. However, until project funds were reduced, we planned to show the relationships between classical control system parameters such as the poles, zeros and

transfer functions to the damping coefficients and frequency response quantities referred to by biomedical researchers. Then we planned to incorporate adaptive equalization techniques to optimally compensate the measurement system's instantaneous frequency response to obtain a flat magnitude, linear phase-shifted measurement system. Experiments can then be performed to compare the clinical studies with the theoretical model. In this manner, we had hoped to develop the mathematical foundation and experimental evidence to suggest a universal standard technique for the comparison of direct and indirect blood pressure measurements.

III. Data Reduction/Data Compression System

Data compression algorithms that were considered include the fan method, Walsh and Fourier Transforms, and interpolation and decimation in time (Caceres, 1965). The linear fan method offers a potential of approximately a 16 to 1 reduction in data with a small computational complexity. Computation of the fast Fourier transform requires on the order of N (log N) complex multiplications and additions; the fast Walsh transform is about three times faster than an efficient fast Fourier transform. The algorithm for the fast Walsh transform contains no multiplications and only N (log N) real additions.

Since funding was stopped halfway through this project, this task was not completed. However, we propose that future studies into data compression/data redundancy algorithms for physiological monitoring continue our lead. Future algorithms must develop a measure of diagnostic value, in conjunction with physiologists, for the comparison of the computerized diagnostic tests versus the resultant compressed waveform.

Many researchers have used the mean square error (mse) as the criterion for comparing their approximations to the original signal. However, in a recent report by Kuklinski and Zied (Kuklinski, 1981),

> "What validity does the mean square error have in terms of being useful measure of the diagnostic content of the reconstructed waveform? From contacts with Air Force and other cardiologists it was apparent that two reconstructions of an ECG signal with the same mean square error resulted in different diagnoses. A more meaningful criterion for evaluating the utility of a reconstructed ECG waveform should be established. This criterion should be based on the measures used by cardiologists in their evaluation of actual ECG data."

The method we propose to evaluate the overall effectiveness of data compression/data reduction experiments is to look at the combined data compression/data reduction system and compare the results of computerized diagnostic tests between the original and compressed waveforms. The figure-of-merit will be weighted by diagnostic accuracy and computational complexity.

We propose a data reduction scheme to store an average waveform and then using template matching or other pattern recognition algorithms, compare each successive waveform to the average waveform and store parameters which relate the nearness of fit. In this manner, we only have to store a data compressed version of abnormal waveforms, and a few bits of information about other waveforms.

IV. Segmented Filter Bank

Advanced signal processing techniques planned for incorporation in physiological flow measuring systems include consideration of a segmented filter bank for

improvement of the signal to noise ratio (SNR) in doppler audio signals. By making some standard assumptions, it can be easily shown that passing a signal (e.g. the Doppler audio signal) through a bank of filters resuls in an improvement in the SNR. This improvement is directly proportional to the square root of the number of filter sections. The application of this technique should provide researchers with a significant improvement in the accuracy of their Doppler ultrasound data when a Zero Crossing Counter is used. The technique seems especially promising for CW systems and could readily be expanded for pulsed Doppler applications.

Denoting the power spectrum of the input noise by $P_{xx}(\omega)$, the total noise power prior to filtering is given by

$$P_{in} = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_{xx}(\omega) d\omega \tag{1}$$

which at the ouput of the filter is

$$P_{out} = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_{xx}(\omega) \quad |H(j\omega)|^2 d\omega \tag{2}$$

where H ($j\omega$) is the frequency response of that filter. Assuming an ideal low pass filter of height A and cut-off frequency ω_c, then equation (2) becomes

$$P_{out} = \frac{A^2}{2} \int_{-\omega_c}^{\omega_c} P_{xx}(\omega) d\omega \tag{3}$$

We can express the improvement in SNR due to the filter in terms of peak signal and R.M.S. noise;

$$SNR_{improvement} = \frac{\text{Peak Signal Out}}{\text{Peak Signal In}} \cdot \frac{\text{RMS Noise In}}{\text{RMS Noise Out}}$$

The ratio R, of noise power at input and output of the filter is eqn (1) divided by eqn (3);

$$R = \frac{\int_{-\infty}^{\infty} P_{xx}(\omega)d\omega}{A^2 \int_{-\omega_c}^{\omega_c} P_{xx}(\omega)d\omega} \qquad (4)$$

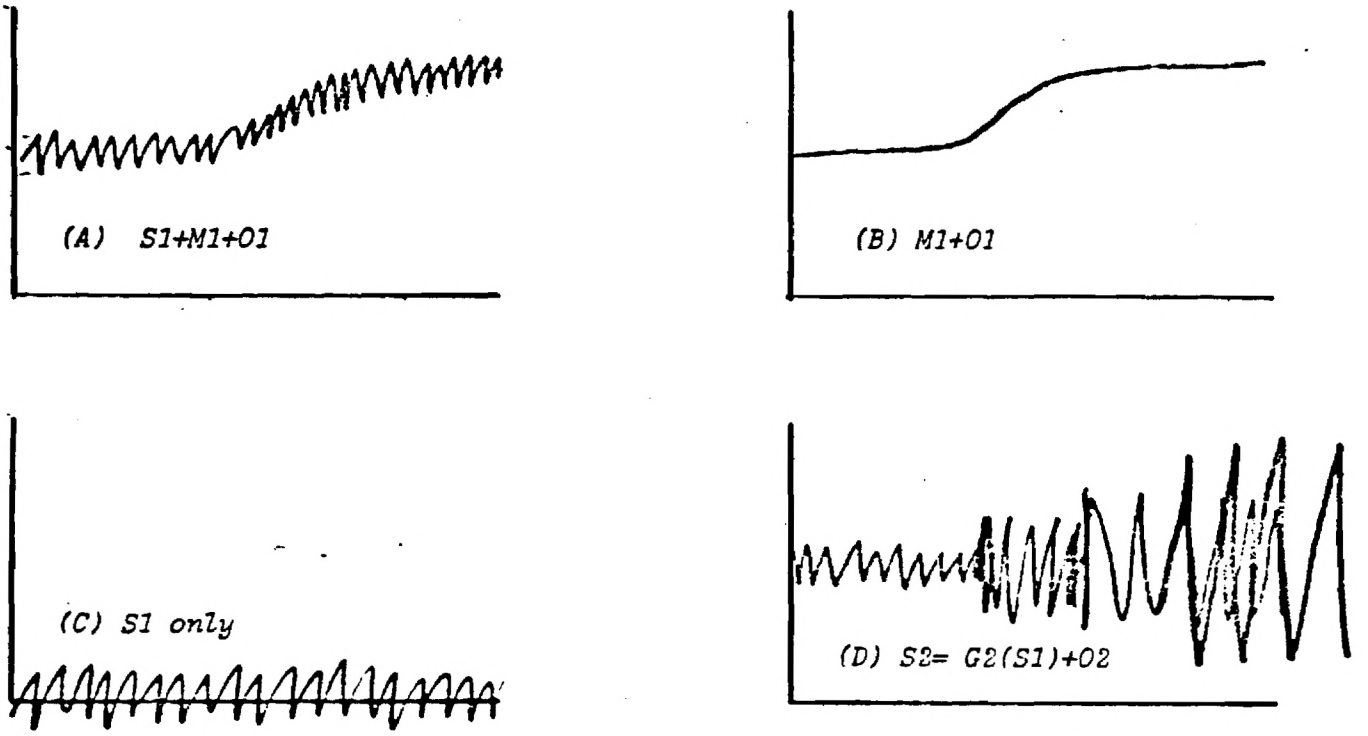
The ratio of noise amplitude, measured in terms of the square root of its mean square value, will therefore be $\sqrt{R}$. The signal amplitude at the filter output will be A times that at the input, therefore, the SNR Improvement is given by:

$$SNR_{improvement} = \frac{\left[\int_{-\infty}^{\infty} P_{xx}(\omega)d\omega\right]^{\frac{1}{2}}}{\int_{-\omega_c}^{\omega_c} P_{xx}(\omega)d\omega}$$

Making a standard assumption that the power spectrum of the noise signal is flat in the range $-\omega_n < \omega < \omega_n$, then

$$\text{SNR IMPROVEMENT} = \left[\omega_n/\omega_c\right]^{\frac{1}{2}}$$

Therefore, if we divide the orignal signal bandwidth, ω_c, into N identical sections, the improvement in the SNR is $\sqrt{N}$.



WAVEFORMS AT VARIOUS POINTS IN MEAN CANCELLATION CIRCUIT

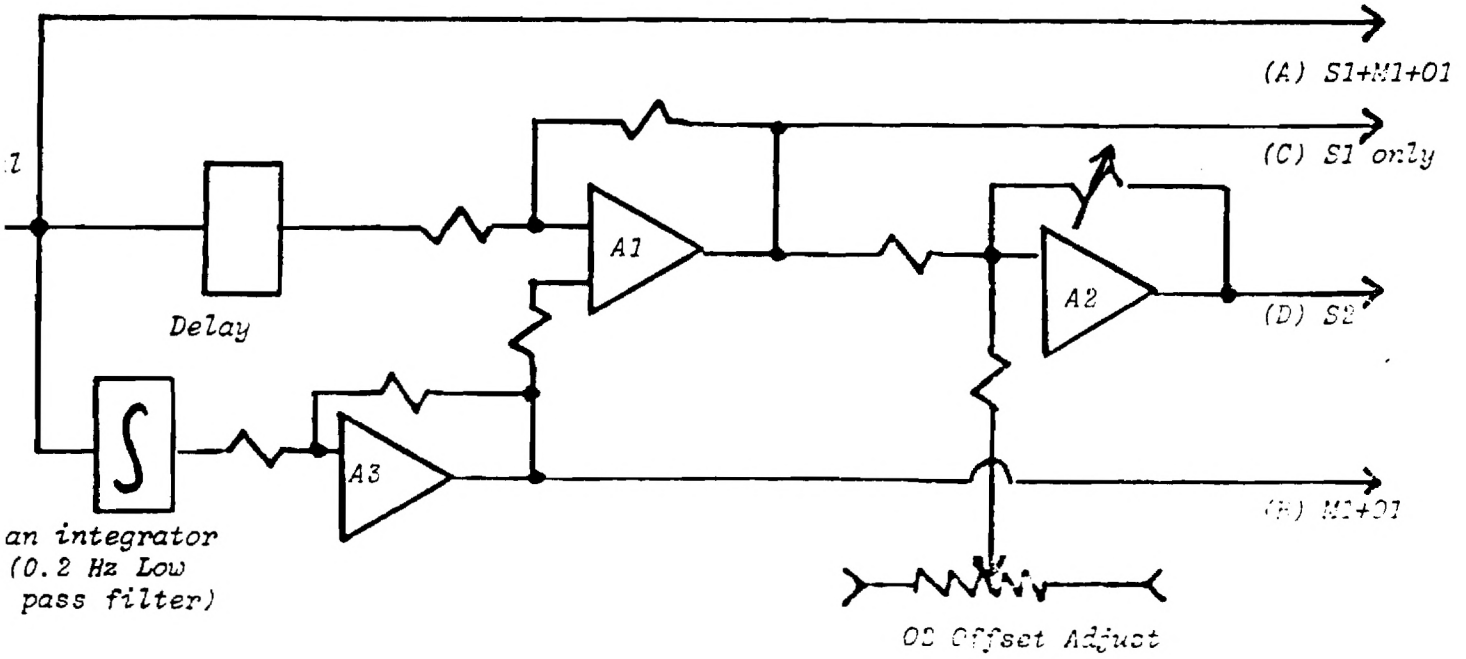


Figure 1. Mean cancellation circuit.

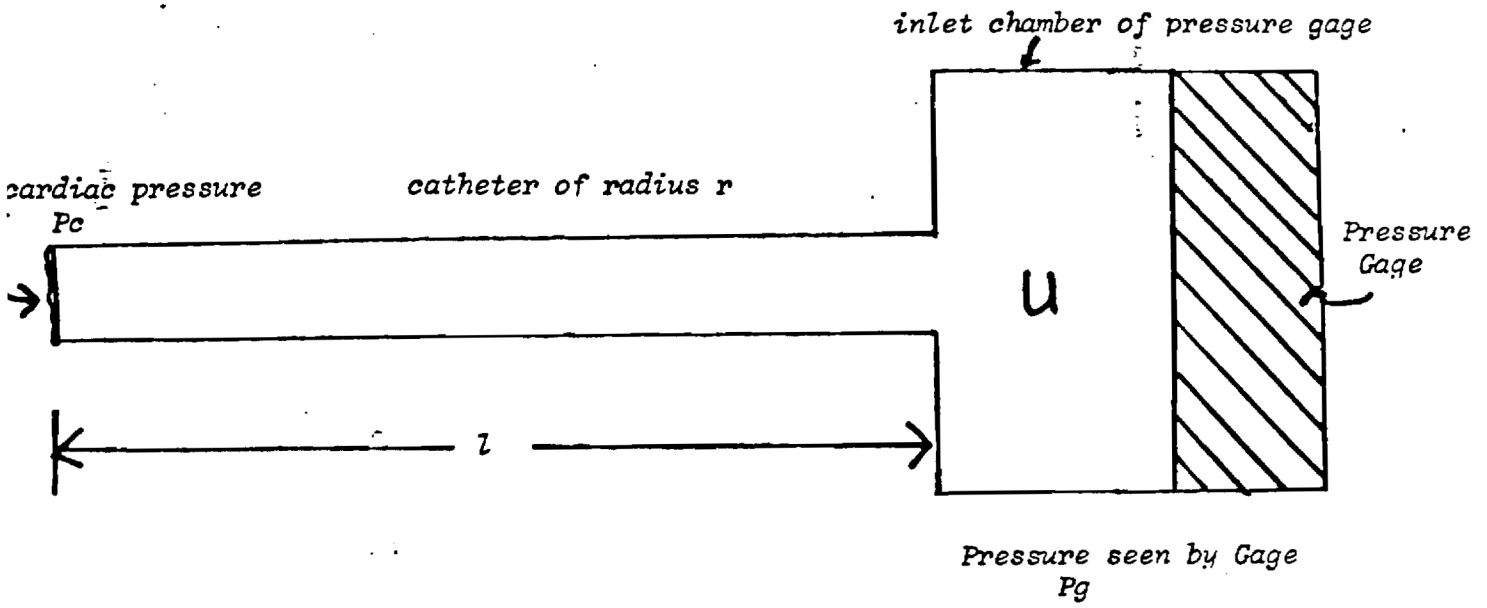


Figure 2. Schematic representation of fluid-filled catheter pressure measurement system (from Welkowitz and Deutsch, 1976).

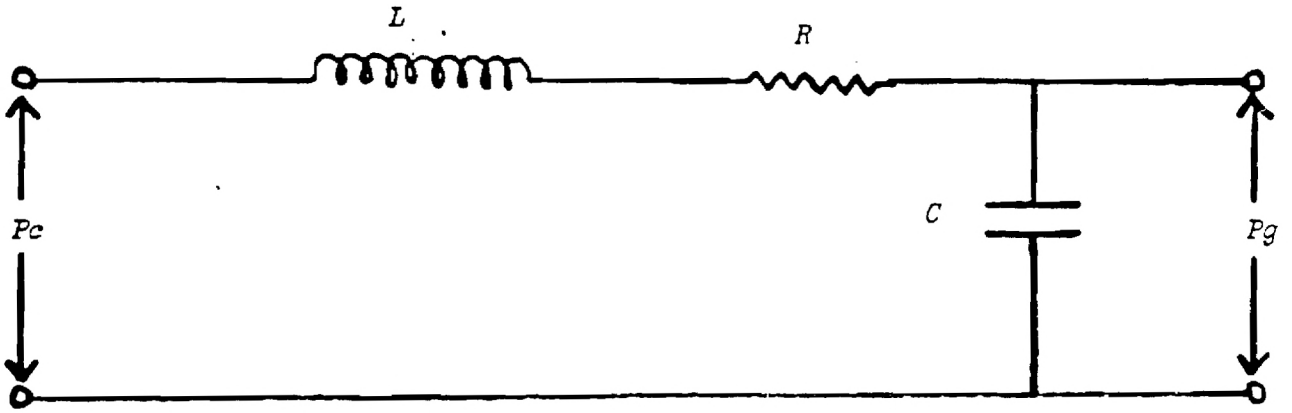


Figure 3. Lumped circuit analogy of fluid-filled catheter.

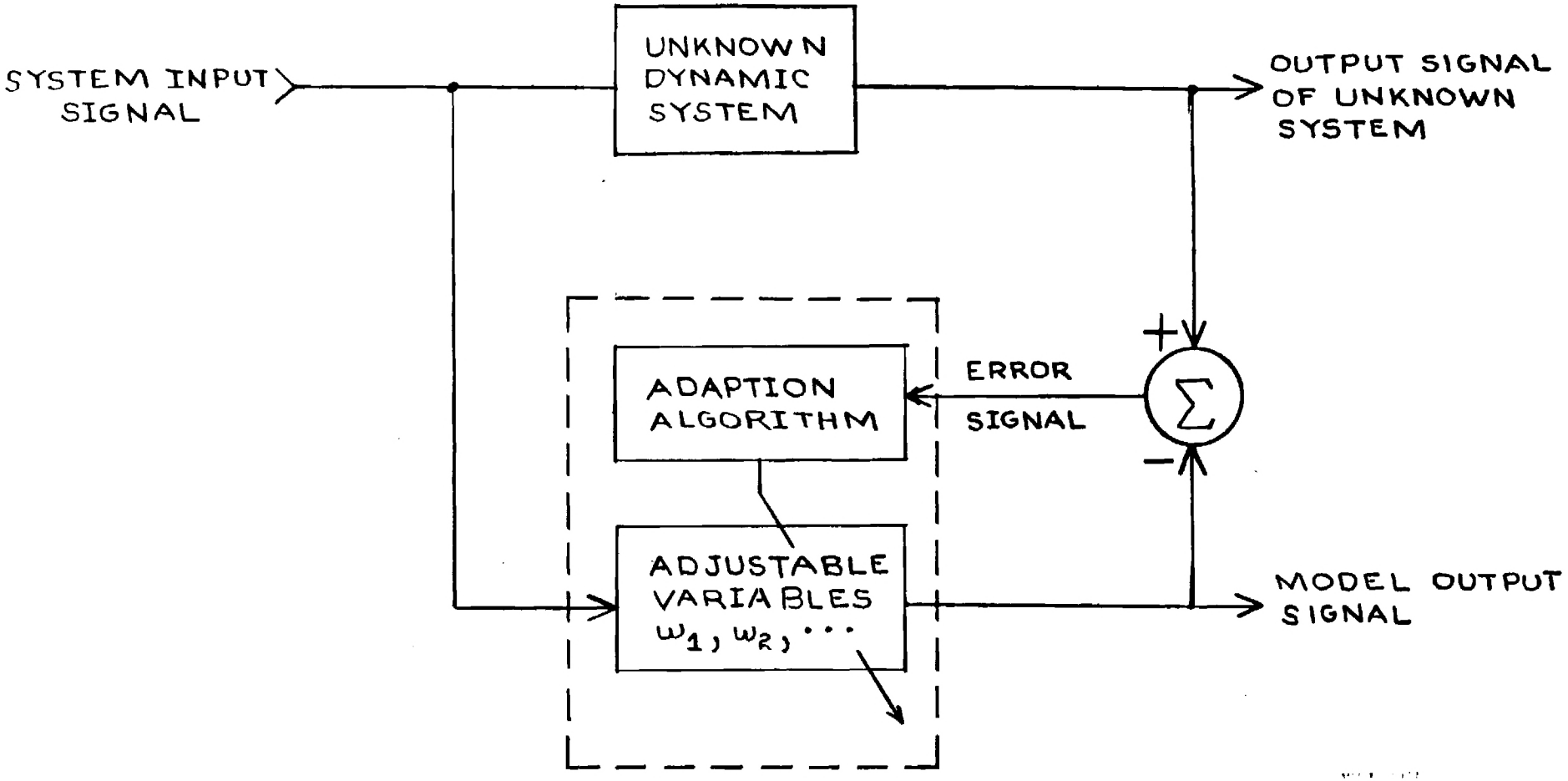


Figure 4. An adaptive modeling system for in-vivo blood pressure measurements

BIBLIOGRAPHY

1. Best and Taylor, Physiological Basis of Medical Practice, 9th Edition.

2. G. D. Zuidema, R. Edelberg, and E. W. Salzman, "A Device for the Indirect Recording of Blood Pressure," Journal of Applied Physiology," 1956.

3. C. R. Smith and W. H. Bickley, The Measurement of Blood Pressure in the Human Body, National Aeronautics and Space Administration Technology Survey, 1964.

4. J. M. R. Bruner, et al, "Comparison of Direct and Indirect Methods of Measuring Arterial Blood Pressure, parts I-III," Medical Instrumentation, January-April, 1981.

5. D. A. Paulus, "Noninvasive Blood Pressure Measurement," Medical Instrumentation, vol. 15, No. 2, March-April, 1981, pp 91-94.

6. J. R. Cox, et al, "AZTEC, A Preprocessing Program for Real-Time ECG Rythm Analysis," IEEE Trans. on Bio-Med Engr, April, 1968, pp 128-129.

7. J. Woodcock, et al, "Physical Aspects of Blood-Velocity Measurements by Doppler-Shifted Ultrasound."

8. D. M. DiPietro and J. D. Meindl, "Optimal System Design for an Implantable CW Doppler Ultrasonic Flowmeter," IEEE Trans on Bio-Med Engr, vol. BME-25, no. 3, May 1978, pp 255-263.

9. W. S. Kuklinski and A. M. Zied, A Fast Walsh Transform Electrocardiogram Data Compression Algorithm Suitable for Microprocessor Implementation USAF School of Aerospace Medicine Report No SAM-TR-81-13, 1980.

10. L. W. Gardenhire, "Data Reduction for Biomedical Telemetry."

11. W. Welkowitz and S. Deutsch, Biomedical Instruments: Theory and Design, Academic Press, New York, 1976.

APPENDIX

ADAPTIVE FILTERS

I. INTRODUCTION

This technical memo presents a tuitorial review of the principles of adaptive filtering. An adaptive filter is a system which is in some sense self-designing, actually self-optimizing. It designs its implulse response by adjusting its internal settings based upon estimated (measured) statistical characteristics of its input signals. The statistics are not measured explicitly and then used to design the filter; rather, the filter design is accomplished automatically by a recursive algorithm that automatically updates the system adjustments with the arrival of each new data sample. This approach does not require complete a priori knowledge of the statistics of the signals to be filtered. Applications of adaptive filtering exist in the fields of noise filtering, automatic control, spectral estimation, pattern recognition, adaptive antenna design, and many others. Unless otherwise noted, the analysis presented below follows the treatment of this subject given by Widrow [1,3,12].

Inevitable errors in the statistical estimates prevent the adaptive filter from delivering optimal performance, but the loss in performance can often be made quite small. This loss is related to the averaging time (which in turn is related to the speed of adaption) and to the number of internal adjustments.

Signal estimation theory describes the estimation of a signal from measurement data corrupted by noise. The optimal estimate of the signal is

one that minimizes the estimation error in some well defined statistical sense. The form of adaptive filter described in this paper is almost as simple to implement as the Wiener filter, and should perform nearly as well as the Kalman-Bucy filter (given complete a priori information). A Wiener filter is the optimum linear filter in the mean-square-error sense through which data samples are passed to obtain the "best" estimate of the actual input signal. Later work by Kalman and Bucy [2] led to the design of optimal time-variable linear filters for nonstationary signals. For such signals, Kalman-Bucy filters can deliver substantially better performance than Wiener filters. When a priori information is not perfectly known, it is quite possible that the peformance of an adaptive filter could exceed that of either a Wiener or a Kalman-Bucy filter. When almost no a priori information is available, the use of an adaptive filter may be the only reasonable way.

II. AN ADAPTIVE FILTER STRUCTURE

Adaptive filters may be continuous or discrete. This technical memo will describe a discrete type of adaptive filter shown in Figure 1. It consists of a tapped delay line, variable weights (variable gains) whose input signals are the signals at the taps of the delay line, a summer to add the weighted signals, and a control processor to automatically adjust the weights. The impulse response of the system is completely controlled by the weight settings. The adaption process automatically seeks an optimal filter impulse response by adjusting the weights.

III. THE PERFORMANCE SURFACE

The principal component of most adaptive systems is the adaptive linear combiner, shown in Figure 2. This combinatorial system can be used in adaptive Wiener filters, adaptive antenna arrays, adaptive control systems, adaptive modeling and estimation. When its output is applied to a threshold device, it can be used as a trainable decision-maker, a pattern classifier, a threshold logic element, or it can be used as the adaptive portion of certain learning control systems [3-6]. The adaptive linear combiner multiplies a set of stationary input signals by the corresponding weight values and sums these products to form an output signal. The input signals in the set are assumed to occur simultaneously and discretely in time. The jth set of input signals is designated by the vector

$$\underline{X}^T(j) = [x_1(j), x_2(j), \cdots, x_n(j)] \ .$$

The set of weights is designated by the vector

$$\underline{W}^T(j) = [w_1(j), w_2(j), \cdots, w_n(j)] \ .$$

The jth output signal, for a fixed set of weights, is

$$y(j) = \sum_{i=1}^{n} w_i(j) x_i(j) \ . \qquad (1)$$

This can be written in matrix form as

$$y(j) = W^T(j)X(j) = X^T(j)W(j) \qquad (2)$$

Two kinds of processes take place in the adpative filter: training and operating. The training (adaption) process is concerned with adjusting the weights. The operating process consists in forming output signals by weighting the input signals, using the weights resulting from the training process.

During the training process, an additional input signal, the "desired response" or training signal, d(j), must be supplied to the adaptive filter along with the usual input signals. This requirement may in some cases restrict the use of this form of an adaptive filter. In most applications some ingenuity is required to obtain a suitable input for d_j. After all, if the desired response were known, why would one need an adaptive processor?

The error signal ε_j is defined as the difference between the desired response, d_j and the adaptive filter output, y_j:

$$\varepsilon_j = d_j - X_j^T W = d_j - W^T X_j \ . \qquad (3)$$

The square of the error is

$$\varepsilon^2(j) = d_j^2 + W^T X_j X_j^T W - 2d_j X_j^T W \ . \qquad (4)$$

Assuming that the input signals are statistically stationary, taking the expected value of both sides gives the mean-square error

$$E[\varepsilon^2(j)] = E[d^2(j)] + W^T E[X(j)X^T(j)]W - 2E[d(j)X^T(j)]W$$

$$E[\varepsilon^2(j)] = E[d^2(j)] + W^T R(x,x)W - 2P(x,d)W \qquad (5)$$

where the cross-correlation vector between the input signals and the desired response is defined as

$$E[d(j)X(j)] = E\begin{bmatrix} x_1(j)d(j) \\ x_2(j)d(j) \\ \vdots \\ x_n(j)d(j) \end{bmatrix} = P(x,d) \qquad (6)$$

and where the correlation matrix of the input signals is defined as

$$E[X(j)X^T(j)] = E\begin{bmatrix} x_1(j)x_1(j) & x_1(j)x_2(j) & \cdots \\ x_2(j)x_1(j) & x_2(j)x_2(j) & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \cdots & x_n(j)x_n(j) \end{bmatrix}$$

$$= R(x,x) \qquad (7)$$

From Equation (5), it can be observed that for stationary input signals, the mean-square error is precisely a second-order (quadratic) function of the internally controllable variables, the weights. The mean-square-error performance function may be visualized as a bowl-shaped surface, a parabolic function of weight variables (a hyperparaboloid if there are more than two weights). See Figure 3. Any quadratic performance surface has a unique minimum point that can be sought by a gradient-based algorithm. The adaptive

process has the job of continually seeking the "bottom of the bowl". It is easiest conceptually to first consider the case when the statistics describing the signal environment are perfectly known. Then the gradient at any point on the performance surface can be determined exactly and the method of steepest descent [7,8] can be employed to find the bottom of the bowl, i.e., the point in space representing the optimal solution, the minimum mean square error.

IV. THE METHOD OF STEEPEST DESCENT

The metod of steepest descent begins with an initial guess of where the performance minimum point may be. This initial guess consists of a set of initial values for each weight of the weight vector components. Having selected a starting point, the gradient vector is then determined, and the next guess is obtained by making an appropriate change in the current guess. This appropriate change is determined by perturbing the weight vector in the opposite direction from the gradient (that is, in the direction of the steepest downward slope of the wall of the bowl-shaped performance surface).

The gradient at any point on the performance surface may be obtained by differentiating the mean-square-error function of equation (5) with respect to the weight vector. The gradient is

$$\nabla[\varepsilon^2(j)] = -2P(x,d) + 2R(x,x)W \ . \qquad (8)$$

To find the "optimal" weight vector, W^*, that yields the least mean-square error, set the gradient to zero. Accordingly,

$$P(x,d) = R(x,x)W^*$$

$$W^* = R^{-1}(x,x)P(x,d) \quad . \tag{9}$$

Equation (9) is the Wiener-Hopf equation in matrix form. The weight vector, W^*, is often referred to as the Wiener weight vector.

In seeking the minimum mean-square error by the method of steepest descent, one begins with an initial guess as to where the minimum point of mean-square-error may be. This means that one begins with a set of initial conditions for the weights. The gradient vector is then measured, and the next guess is obtained from the present guess by making a change in the weight vector in the direction of the negative of the gradient vector - that is, in the opposite direction of the gradient vector, the process will converge on the stationary (minimum) point regardless of the choice of initial weights. The weights undergo geometric (discrete exponential) transients in relaxing toward the surface minimum. The amount of damping can be controlled by a scalar variable μ.

The method of steepest descent makes each change in the weight vector proportional to the gradient vector; the method of steepest descent can be described by the following relation:

$$\underline{W}(j+1) = \underline{W}(j) + \mu(-\underline{\nabla}) \quad . \tag{10}$$

The gradient vector $\nabla[\varepsilon(j)]$ can be obtained by using equation (8), therefore,

$$W(j+1) = W(j) - 2\mu R(x,x)W(j) + 2\mu P(x,d) \quad . \qquad (11)$$

The gradient vector $\nabla[\varepsilon^2(j)]$ is the gradient of the expected error-squared function when the weight vector is W(j).

When, as in the present case, the performance function is quadratic, the gradient is a linear function of the weights.

Equation (11) can be rewritten as follows:

$$W_{j+1} + (2\mu R - I)W_j = 2\mu RW^* \quad . \qquad (11a)$$

This equation is a linear, multivariable, crosscoupled first order difference equation. In the next section a method will be shown to uncouple this equation and in so doing display geometric properties of the quadratic performance surface.

V. GEOMETRIC PROPERTIES OF THE QUADRATIC PERFORMANCE SURFACE

This section will present a geometric representation of the transient analysis of the quadratic performance surface in terms of the eigenvalues and eigenvectors of the input autocorrelation matrix.

An expression for the minimum mean-square error may be obtained by substituting (9) into (5):

$$J^* = E[\varepsilon^2_{min}] = E[d^2(j)] - W^{*T}P(x,d) \quad . \qquad (12)$$

In order to facilitate a transient analysis, we can define a difference vector

$$V = W - W^* \tag{13}$$

to be the deviation of the weight vector from the optimum Wiener weight vector.

Combining this with (5), (9) and (12) yields

$$\begin{aligned} J &= E[d_j^2] - P^T W - W^T P + W^T R W \\ &= E[d_j^2] - P^T(W^*+V) - (W^*+V)^T P + (W^*+V)^T R(W^*+V) \\ &= J^* - P^T V - W^{*T} P - V^T P + W^{*T} R W^* + W^{*T} R V + V^T R W^* + V^T R V \\ &= J^* + V^T R V \\ &= J^* + (W-W^*)^T R(W-W^*) \end{aligned} \tag{14}$$

The matrix R is the input autocorrelation matrix. The main diagonal terms are the mean squares of the input-signal components, and the cross terms are crosscorrelations among the inputs. The autocorrelation function of a signal x_k is defined as the expected value of the product of x_k and its own replica shifted by ℓ time intervals. Therefore,

$$r_\ell = E[x_k x_{k-\ell}] = E[x_{k+\ell} x_k]$$

for stationary sequences, and therefore,

$$E[x_{k+\ell}x_k] = E[x_k x_{k+\ell}]$$

$$r_\ell = r_{-\ell} .$$

This is the symmetry property of the autocorrelation function. Therefore, R is a square symmetric matrix. And since R is symmetric, the product $V^T RV$ is a pure quadratic form. Since x_k are physically real signals, R is also real and the mean square error has to be greater than or equal to zero. Therefore R is also positive definite or at least positive semidefinite. Note that (14) is a quadratic form in W whose value is the minimum mean square error when $W = W^*$.

It is also clear from (14) that the orientation and shape of the quadratic mean square error performance surface is a function of the input signal autocorrelation matrix, R. Geometric properties of the shape of the surface can be determined by expressing R in normal form, in terms of its eigenvalues and eigenvectors.

The eigenvalues of R are obtained from the characteristic equation of R,

$$\det [R - \lambda I] = 0 . \qquad (15a)$$

The N solutions of (15a) are the eigenvalues of $R, \lambda_1, \lambda_2, \cdots, \lambda_n$. Corresponding to each eigenvalue of R there exists at least one eigenvector, q_p, of R from the equation

$$Rq_p = \lambda_p q_p . \qquad (15b)$$

The vector q_p is the p^{th} eigenvector of R. Extending (15b) yields

$$R[q_0,q_1,\cdots q_n] = [q_0,q_1,\cdots q_n] \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \qquad (16)$$

Equation (16) can be written as

$$RQ = Q\Lambda \, . \qquad (17)$$

The eigenvalue matrix, Λ, is diagonal. All of its elements are zero except for the main diagonal whose elements are the set of eigenvalues of R.

The matrix Q is called the modal matrix of R. Its columns are the eigenvectors of R. Both Λ and Q are square with dimensions N × N, just like R.

By the definition of R, it is apparent that R is a symmetric matrix, $R = R^T$. Since R is symmetric, its eigenvectors corresponding to distinct eigenvalues must be orthogonal. Since R is real (all of its elements are real numbers) in addition to being symmetric, all of its eigenvalues must be real.

The columns of the modal matrix, Q, are assumed to be orthonormal. Eigenvectors of distinct eigenvalues are orthogonal. If repeated eigenvalues occur, the corresponding eigenvectors can be chosen so they too are orthogonal. Thus, all N eigenvectors of R, i.e., the columns of Q, can be chosen to be mutually orthogonal and normalized. Q is therefore orthonormal. Consequently,

$$QQ^T = I \ . \tag{18}$$

Therefore,

$$Q^{-1} = Q^T \ . \tag{19}$$

Thus the inverse of Q always exists.

Postmultiplying both sides of (17) by Q^{-1} yields an important result

$$R = Q\Lambda Q^T \tag{20}$$

$$= \sum_{p=1}^{n} \lambda_p \, q_p q_p^T$$

by Mercer's Theorem.

In order for the mean-square error to always be positive (as is necessary in a real physical problem), the matrix R must be positive definite or positive semidefinite. Therefore, the eigenvalues of R must be greater than or equal to zero.

Substituting (20) into (14) yields

$$E[\varepsilon^2(j)] = J = J_{min}^{*} + V^T Q \Lambda Q^T V \ . \tag{21}$$

A new set of coordinates can be defined as follows:

$$W' = Q^T W$$

$$W^{*'} = Q^T W^* \quad . \tag{22}$$

Therefore,

$$V' = (W' - W^{*'}) = Q^T(W - W^*) = Q^T V \quad .$$

Thus,

$$V'^T = V^T Q \quad . \tag{23}$$

Substituting (22) and (23) into (21) yields

$$J = J_{min} + V'^T \Lambda V' \quad . \tag{24}$$

Since we can arbitrarily choose V' by choice of $V = QV'$, the only way for J to be greater than or equal to zero for all values of V (and hence V') is for all the eigenvalues of R to be positive (R is positive definite) or for some eigenvalues to be zero and for some to be positive (in which case R is positive semidefinite).

The transformation Q projects W(j) into $W'(j)$, i.e., projects W(j) onto the primed coordinates. From (24) it can be observed that since Λ is diagonal, the primed coordinates must comprise the principal axes of the quadratic mean square error performance surface. The transformation rotates the "natural" coordinates into the "primed" coordinates. No cross terms

appear in (24) because Λ is diagonal. Therefore, the coordinates V' are the set of principal axes of the quadratic performance surface, i.e., the set of eigenvectors of R which define the principal axes of the mean square error performance surface.

The eigenvalues also have important physical significance. Rewriting (14) as

$$J = J_{min} + (W' - W^{*'})^T \Lambda (W' - W^{*'})$$

$$= J_{min} + \sum_{p=1}^{n} \lambda_p (w_p' - w_p^{*'})^2$$

where w_p' , $w_p^{*'}$ are defined as the components of W' and $W^{*'}$ respectively.

For the one-dimensional case the mean square error is given by

$$J = J_{min} + \lambda(w - w^*)^2 \ . \tag{26}$$

The first derivative of J with respect to w is

$$\frac{dJ}{dw} = 2\lambda(w - w^*) \ . \tag{27}$$

The second derivative is

$$\frac{d^2J}{dw^2} = 2\lambda \ . \tag{28}$$

In one dimension, 2λ is the second derivative of the parabolic J surface, being equal to the curvature of the parabola at its minimum. In the

multidimensional case each eigenvalue is the coefficient of the second degree term of J giving the steepness of the surface.

In summary, the eigenvalues of the input autocorrelation matrix R, are equal to half the value of the second partial derivative of the J-surface in the direction of the associated eigenvector, and the eigenvectors determine the directions of the principal axes of the surface.

VI. STABILITY AND RATE OF CONVERGENCE

The iterative convergent process of steepest descent was presented by (10)

$$W_{j+1} = W_j + \mu(-\nabla_j) \ . \qquad (10)$$

The parameter μ controls the convergence rate and stability of the process. The effect of the choice of μ upon the iterative process will be shown below.

The dynamics (transient behavior) of the iterative process, starting from some initial guess, and relaxing toward the optimum setting, W^*, can be analyzed from a study of (11a):

$$W_{j+1} + (2\mu R - I)W_j = 2\mu R\, W^* \ . \qquad (11a)$$

It is necessary to find the general solution to this equation. This is difficult because the various components of W_j are cross-coupled. The matrix coefficient of W_j in (11a) is non-diagonal because it contains the term $2\mu R$. In general, R is not diagonal. To find the general solution to (11a) it is

first necessary to uncouple the components of W_j. Substituting (20) and (22) into (11a) yields a difference equation expressed in the primed coordinates as

$$W'_{j+1} + (2\mu\Lambda - I)W'_j = 2\mu\Lambda W^{*'} . \tag{29}$$

By expressing the dynamics of the iterative process in a difference equation in the primed coordinate system the action of the steepest-descent relaxation process is expressed without cross-coupling, i.e., I and Λ are diagonal.

The general solution is

$$\begin{aligned} W'_j &= W^{*'} + (I - 2\mu\Lambda)^j (W'_o - W^{*'}) \\ V'_j &= (I - 2\mu\Lambda)^j V'_o \end{aligned} \tag{30}$$

where W_o is the initial weight setting. Equation (30) shows that if the method of steepest descent is to be stable (convergent),

$$\lim_{j\to\infty} (I - 2\mu\Lambda)^j = 0 . \tag{31}$$

When this condition is satisfied

$$\lim_{j\to\infty} W'_j = W^{*'}$$

therefore,

$$\lim_{j\to\infty} W_j = W^* . \tag{32}$$

Equation (31) is stable in the one-dimensional case when

$$\left|(1 - 2\mu\lambda)\right| < 1$$

therefore

$$1/\lambda_{max} > \mu > 0 \qquad (33)$$

where λ_{max} is the largest eigenvalue of R. Condition (33) is necessary and sufficient for convergence of the method of steepest descent when applied to the searching of a quadratic surface.

VII. The LMS Algorithm

In most practical situations, the signal statistics are not known *a priori* so the exact gradient cannot be determined. The LMS algorithm is exactly like the method of steepest descent except that the gradient vector is estimated instead of computing the exact gradient. The Least Mean Square Error (LMS) algorithm of Widrow and Hoff [9] is a practical method for finding approximate solutions to equation (9) on a real-time basis. The accuracy of the method is limited by statistical sample size, since the weight values found by the algorithm are based on finite-time measurement of input signals. The algorithm does not require explicit measurements of correlation functions, nor does it require matrix inversion, only multiplication and addition. It is based on gradient-search techniques applied to mean-square-

error functions. The algorithm does not require squaring, averaging or differentiation in order to make use of gradients of mean-square-error functions.

The LMS algorithm is based on the method of steepest descent (10). The true gradient at each iteration is generally unknown. However, estimated gradients can be readily obtained. The LMS algorithm implements the method of steppest descent using estimated gradients. The instantaneous gradient can be estimated in a crude but highly effective way by assuming that a single sample of error squared ϵ_j^2 is an estimate of the mean square error and by differentiating ϵ_j^2 with respect to W. Equations (34) show relationships between true and estimated gradients

$$\nabla_j \triangleq \begin{bmatrix} \frac{\partial E[\epsilon_j^2]}{\partial w_o} \\ \vdots \\ \frac{\partial E[\epsilon_j^2]}{\partial w_N} \end{bmatrix} \quad ; \quad \hat{\nabla}_j = \begin{bmatrix} \frac{\partial \epsilon_j^2}{\partial w_o} \\ \vdots \\ \frac{\partial \epsilon_j^2}{\partial w_N} \end{bmatrix} \qquad (34)$$

The estimated gradient components are related to the partial derivatives of the instantaneous error with respect to the weight components, which can be obtained by differentiating (4). Therefore, the gradient estimate is

$$\hat{\nabla}_j = -2\epsilon_j X_j \, . \qquad (35)$$

Using this gradient in place of the true gradient in (10) yields the Widrow-Hoff LMS algorithm:

$$W_{j+1} = W_j + 2\mu\epsilon_j X_j \, . \qquad (36)$$

It has been shown that the gradient estimate used in the LMS algorithm is unbiased and that the expected value of the weight vector converges in the mean to the Wiener weight vector [10,11] when the input vectors are uncorrelated over time (but they could be correlated from input component to component).

The algorithm will converge in the mean and remain stable as long as the parameter μ is greater than 0 but less than the reciprocal of the largest eigenvalue of the matrix R

$$1/\lambda_{max} > \mu > 0 \ .$$

A conservative stability bound for μ is

$$\frac{1}{\sum_{p=1}^{n} \lambda_p} = \frac{1}{\text{trace } R} > \mu > 0 \qquad (38)$$

where the trace of R is equal to the total input power.

ja

REFERENCES

1. B. Widrow, "Adaptive Filters I: Fundamentals", Stanford Electronic Laboratories Technical Report No. 6764-6, Dec. 1966.

2. Kalman, R. and Bucy, R., "New Results in Linear Filtering and Prediction Theory", Trans. ASME, Ser. D. J. Basic Engr., Vol. 83, Dec. 1961, pp. 95-107.

3. B. Widrow, et al, "Adaptive Noise Cancelling: Principles and Applications", Proc. of the IEEE, Vol. 63, No. 12, Dec. 1975, pp. 1692-1716.

4. J. Triechler, "Transient and Convergent Behavior of the Adaptive Line Enhancer", IEEE Trans. on Acoustics, Speech and Sig. Proc., Vol. ASSP-27, No. 1, Feb. 1979, pp. 53-62.

5. J. R. Zeidler, et al, "Adaptive Enhancement of Multiple Sinusoids in Uncorrelated Noise", IEEE Trans. on Acoustics, Speech and Sig. Proc., Vol. ASSP-26, No. 3, June, 1978, pp. 240-253.

6. J. S. Koford, G. F. Groner, "The Use of an Adpative Threshold Element to Design a Linear Optimal Pattern Classifier", IEEE Trans. on Information Theory, Vol. 12, No. 1, Jan. 1966, pp. 42-50.

7. R. V. Southwell, Relaxation Methods in Engineering Sciences, New York, Oxford, 1940.

8. D. J. Wilde, Optimum Seeking Methods. Englewood Cliffs, N. J., Prentice-Hall, 1964.

9. Widrow, B. and Hoff, M., "Adaptive Switching Circuits, IRE Wescon Conv. Rec., pt. 4, 1960, pp. 96-104.

10. Riegler, R. and Compton, R., "An Adaptive Array for Interference Rejection", Proc. IEEE, Vol. 61, pp. 748-758, June 1973.

11. Widrow, B. et al, "Adaptive Antenna Systems", Proc. IEEE, Vol. 55, pp. 2143-2159, Dec. 1967.

12. Widrow, B., "Adaptive Systems Class Notes 1974", Stanford University.

ja

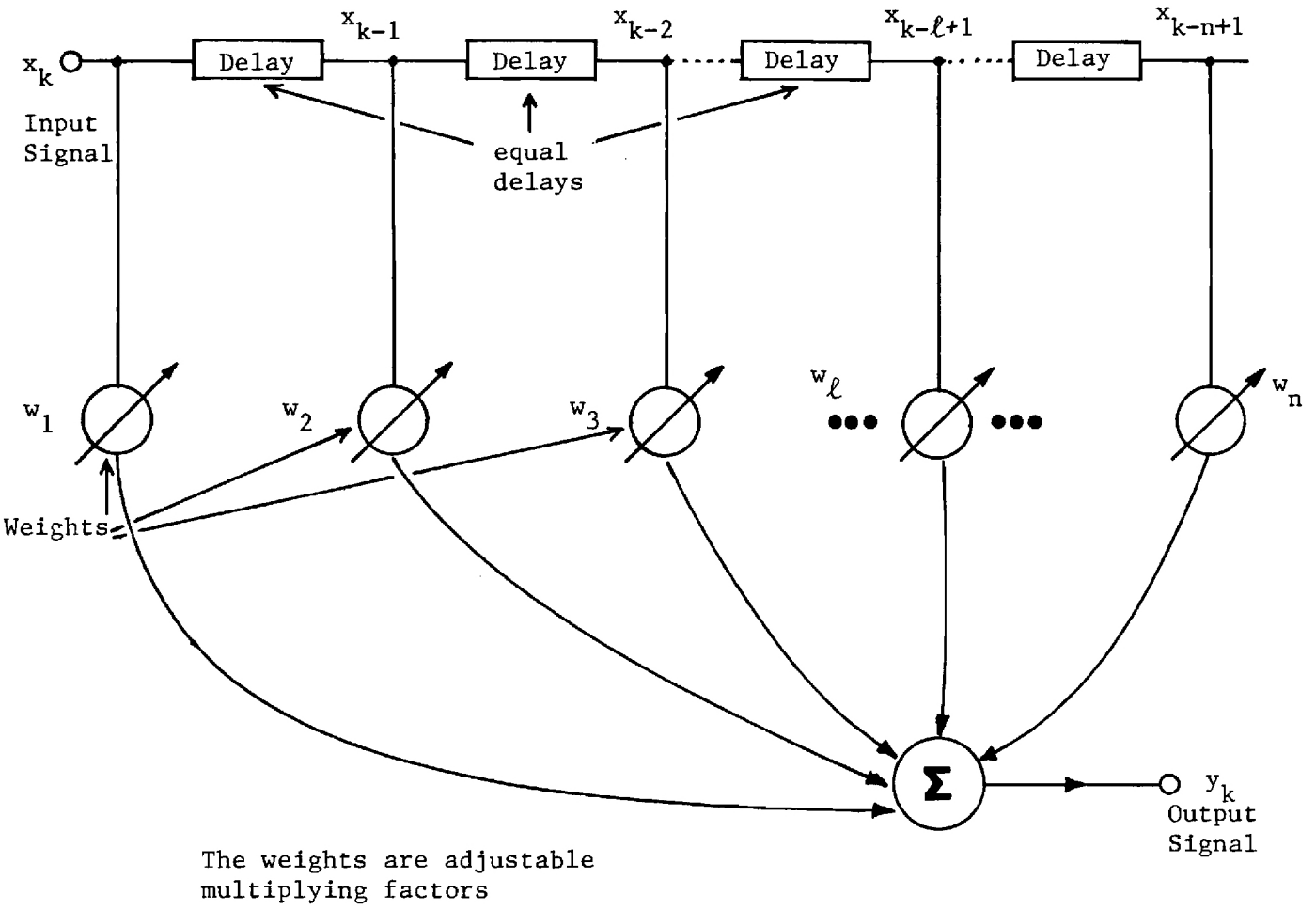


Figure A1. A digital Adaptive filter.

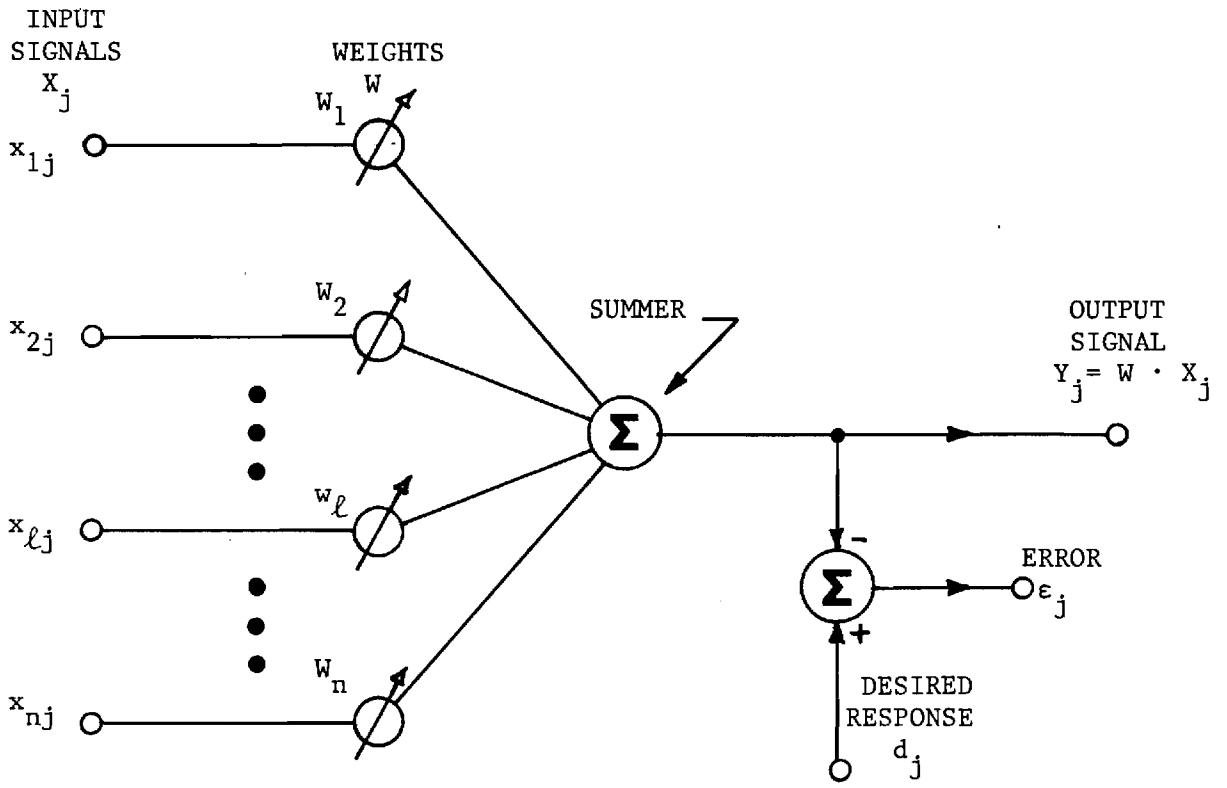


Figure A2. The Adaptive Linear Combiner.

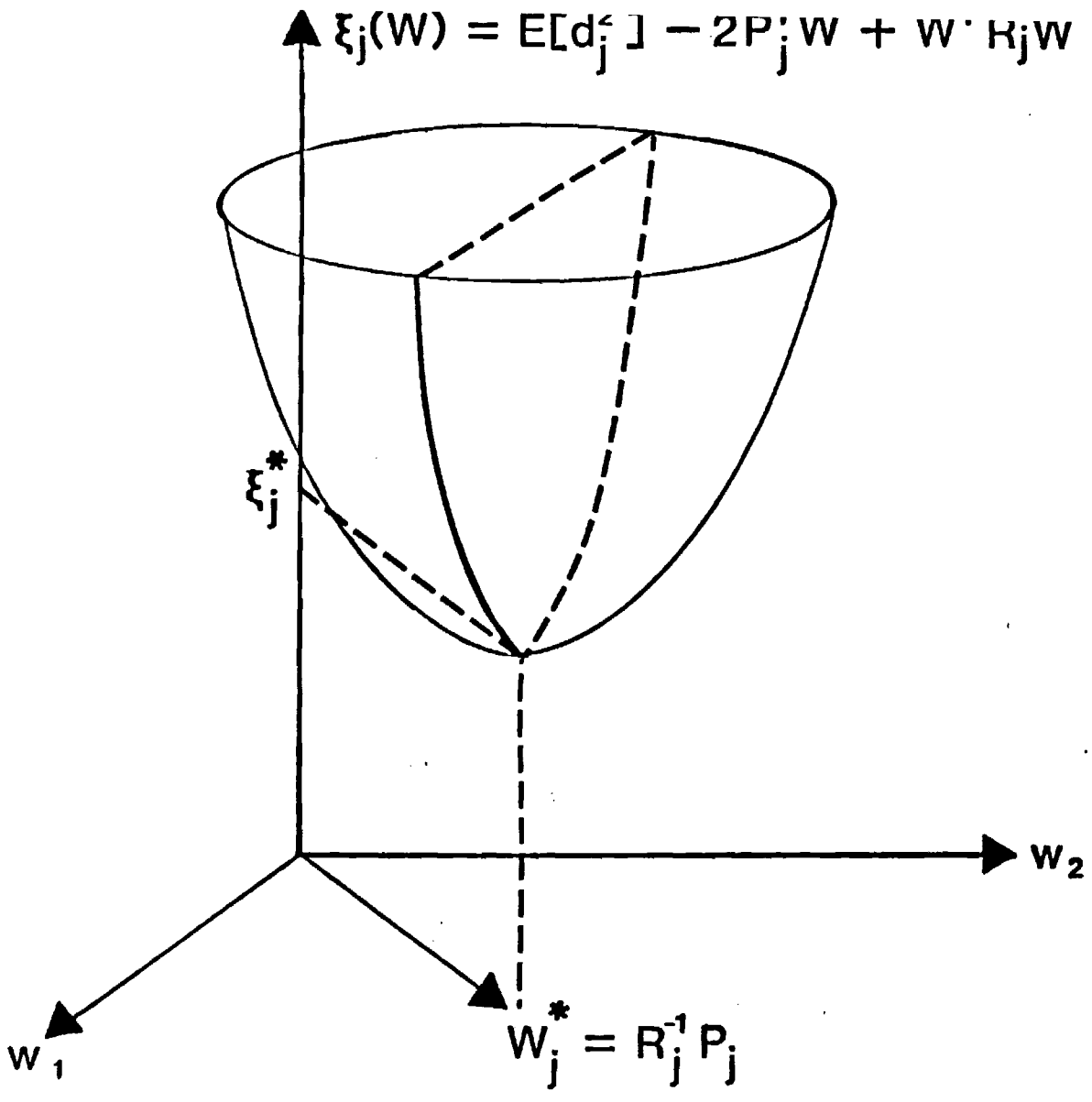


Figure A3. A mean-square-error surface for a two weight filter.